KERNEL REGRESSION ESTIMATION FOR SPATIAL FUNCTIONAL RANDOM VARIABLES

Sophie Dabo-Niang, Mustapha Rachdi, Anne-Françoise Yao

Kernel regression estimation for spatial functional random variables

Sophie Dabo-Niang [♯], Mustapha Rachdi [∓] and Anne-Françoise Yao [§]

Corresponding author : ♯ University Charles De Gaulle, Lille 3, Laboratoire EQUIPPE, maison de la recherche, domaine du pont de bois, BP 60149, 59653 Villeneuve d'ascq cedex, France. sophie.dabo@univ-lille3.fr

§ University Aix-Marseille 2, Laboratoire LMGEM, Campus de Luminy, case 901, 13288 Marseille cedex 09, France. anne-francoise.yao@univmed.fr

∓ University Grenoble 2, UFR SHS, BP. 47 38040 Grenoble Cedex 09 France, Mustapha.Rachdi@upmf-grenoble.fr

Abstract

Given a spatial random process $\left\{(X_{\mathbf{i}},\ Y_{\mathbf{i}}) \in \mathcal{E} \times \mathbb{R},\ \mathbf{i} \in \mathbb{Z}^N\right\}$, we investigate a nonparametric estimate of the conditional expectation of the real random variable $Y_{\mathbf{i}}$ given the functional random field $X_{\mathbf{i}}$ valued in a semi-metric space $\mathcal{E}$. The weak and strong consistencies of the estimate are shown and almost sure rates of convergence are given. Special attention is paid to apply the regression estimate introduced to spatial prediction problems.

Key Words: Regression estimation; Random fields; Functional variables; Infinite dimensional space; Small balls probabilities;.

1. Introduction

Nowadays, the progress of informatics tools permits the recuperation of increasingly bulky data. These large data sets are available essentially by real time monitoring and computers can manage such databases. The object of statistical study can then be curves (consecutive discrete recordings are aggregated and viewed as sampled values of a random curve) not numbers or vectors. Functional data analysis (see Bosq [4], Ferraty and Vieu [18], Ramsay and Silverman [30]) can help to analyze such high-dimensional data sets. The statistical problems involved in modeling functional random variables offer increasing interests in recent literature. For prediction problems, functional data analysis techniques outdo the other approaches because they take advantage of past information. See for example Bosq [4], Besse et al. [2], Cardot and Johannes [8], Fernandez de Castro et al. [17], Ferraty and Vieu [18], Dabo-Niang and Ferraty [11], Dabo-Niang and Rhomari [12], Ferré and Yao [19], Rachdi and Vieu [29] for some regression and forecasts results obtained by several functional models (linear, nonlinear, neural networks and semi-parametric models) for non spatial variables. It is proved in some of these papers that the best predictions were generally obtained by the functional methods (either autoregressive or functional kernel).

To the best of our knowledge, although potential applications of regression estimation (or prediction) to functional spatial data are without number, only the papers of Dabo-Niang and Yao [14], Delicado et al. [15] and Nerini et al. [28], have paid attention to study regression or prediction for functional random fields.

The last two works deal with spatial kriging (linear predictor) methods for spatial functional data. Dabo-Niang and Yao [14] introduced a nonparametric predictor for functional random fields, for which the behavior has not been investigated. We want to go further and extend functional data nonparametric analysis techniques to the spatial domain. We suggest a nonparametric regression estimation approach which is to aggregate over space.That is, we are mainly concerned with kernel regression methods for functional random fields (spatial curves). Note that Lakasaci and Fouzia [23] considered the case of conditional quantile estimation where the regressor take values in a semi-metric space and showed the strong and weak consistency of the conditional quantile.

Spatial regression estimation is an interesting and crucial problem in statistical inference for a number of applications (as in a variety of fields, including soil science, geology, oceanography, econometrics, epidemiology, environmental science, forestry,...), where the influence of a vector of covariates on some response variable is to be studied in a context of spatial dependence.

The literature on parametric spatial modeling is relatively abundant and we refer to Chilès and P. [9], Guyon [20], Anselin and Florax [1], Cressie [10] or Ripley [32] for a list of references. However, the nonparametric treatment of functional or multivariate spatial data is limited. Actually, if some references on spatial nonparametric regression estimation in the multidimensional setting (key refererences are: Lu and Chen [25, 26], Hallin et al. [22], Hallin and Yu [21], Carbon et al. [5], Dabo-Niang and Yao [14], Li and Tran [24],...) already exist, no work have been devoted to nonparametric regression for functionnal spatial data. The aim of this paper is to study the behavior of the functional spatial counterpart of the Nadaraya-Watson Nadaraya [27], Watson [34]'s estimator.

The paper is organized as follows. In Section 2, we provide the notations, assumptions and introduce a kernel estimate of the conditional mean. Section 3 is devoted to convergence in probability and strong convergence of the kernel estimate. To check the performance of our estimator, simulations results will be given in Section 4. Conclusion is given on Section 5 and Proofs and technical lemmas are given in the last section.

2. General setting

Let $\mathbb{Z}^N$ $(N \geq 1)$, be the integer lattice points in the $N-$dimensional Euclidean space. A point in bold $\mathbf{i} = (i_1, ..., i_N) \in \mathbb{Z}^N$ will be referred as a *site*. In fact, spatial data can be seen as realizations of a measurable strictly stationary spatial process Z_i, defined on a probability space $(\Omega, \ \mathcal{A}, \mathbf{P})$. In this paper we deal with process of the feature: $Z_{\mathbf{i}} = (X_{\mathbf{i}}, Y_{\mathbf{i}})$, $\mathbf{i} \in \mathbb{Z}^N$ such that: the $Z_{\mathbf{i}}'$s have the same distribution as a variable $Z = (X, Y)$, where Y is a real-valued and integrable variable and X valued in a separable semi-metric space $(\mathcal{E}, d(.,.))$ (of eventually infinite dimension).

In the following, we will denote by μ the probability distribution of X and $\forall\, \mathbf{i}, \mathbf{j}$ by $\nu_{\mathbf{i},\mathbf{j}}$ the joint probability distribution of $(X_{\mathbf{i}}, X_{\mathbf{j}})$, $\|.\|$ will denote any norm over $\mathbb{Z}^N$, C an arbitrary constant and $B(x, \ \rho)$ the opened ball of center x and radius ρ. We will write $\mathbf{n} \to +\infty$ if $\min_{k=1,\dots,N} n_k \to +\infty$ and we set $\widehat{\mathbf{n}} = n_1 \times ... \times n_N$ and $\mathbb{N}^N$, will denotes subspace of $\mathbb{Z}^N$ of vectors with non-negative components.

For a seek of simplicity, we will suppose that the variable Y is bounded. Instead of this condition one can assume (as in Tran [33], Dabo-Niang and Yao [14] and Ferraty and Vieu [18]) that one of the following assumptions holds:

- For all $\mathbf{i} \neq \mathbf{j}$, $E\left[\mathrm{Y}_{\mathbf{i}}\mathrm{Y}_{\mathbf{j}} | (X_{\mathbf{i}}, X_{\mathbf{j}})\right] \leq C$ for some constant $C > 0$.
- For all $m \geq 2$, $E\left[Y^m | X = x\right] \leq g_m(x) < \infty$, where g_m is a continuous function at $x \in \mathcal{E}$.

The proofs given here remain valid in one of these above cases.

2.1. The spatial kernel regression estimator for functional data

We aim to estimate the regression function $r(x) = E\left(Y | X = x\right)$ of Y given X. To do this, we propose the kernel estimate of the function r based on observations of the process $(Z_{\mathbf{i}})$ in some region $\mathcal{I}_{\mathbf{n}}$ and without the lost of generality, we suppose that $\mathcal{I}_{\mathbf{n}}$ is a rectangular region $\mathcal{I}_{\mathbf{n}} := \left\{\mathbf{i} \in \mathbb{N}^N \ : \ 1 \leq i_k \leq n_k, \ k = 1, ..., N\right\}$:

$$r_{\mathbf{n}}(x) = \begin{cases} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i}} W_{\mathbf{i},x} & \text{if } \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} W_{\mathbf{i},x} \neq 0; \\ \frac{1}{\widehat{\mathbf{n}}} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i}} & \text{else.} \end{cases}$$

where

$$W_{\mathbf{i},x} = \frac{K\left(d(X_{\mathbf{i}}, x) h_{\mathbf{n}}^{-1}\right)}{\sum_{\mathbf{j} \in \mathcal{I}_{\mathbf{n}}} K\left(d(X_{\mathbf{j}}, x) h_{\mathbf{n}}^{-1}\right)}.$$

This estimate can also be written as follows:

$$r_{\mathbf{n}}(x)=\begin{cases}\varphi_{\mathbf{n}}(x)/f_{\mathbf{n}}(x) & \text{if } \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} W_{\mathbf{i},x}\neq 0;\\ \frac{1}{\widehat{\mathbf{n}}}\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i}} & \text{else.}\end{cases}$$

with

$$\varphi_{\mathbf{n}}(x)=\frac{1}{\widehat{\mathbf{n}}E\left(K\left(d(X,x)h_{\mathbf{n}}^{-1}\right)\right)}\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i}}K\left(d(X_{\mathbf{i}},x)h_{\mathbf{n}}^{-1}\right),\ x\in\mathcal{E},$$

$$f_{\mathbf{n}}(x)=\frac{1}{\widehat{\mathbf{n}}E\left(K\left(d(X,x)h_{\mathbf{n}}^{-1}\right)\right)}\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} K\left(d(X_{\mathbf{i}},x)h_{\mathbf{n}}^{-1}\right),\qquad x\in\mathcal{E},$$

where $\lim_{\mathbf{n}\to+\infty} h_{\mathbf{n}}=0\,(+)$, the kernel K is a function from $\mathbb{R}^+$ to $\mathbb{R}^+$.

That is an extension of the well known of Nadaraya [27], Watson [34]'s estimate introduced for *i.i.d* observations.

Since we are dealing with spatial data, as any spatial model, ours must take into account the dependance between observations. Let us now consider some measures of dependence.

2.2. Spatial dependence measures

As it often occurs in spatial dependent data analysis, one needs to defined the type of dependence. We consider here the following two dependence measures:

2.2.1. Local dependence condition

We assume also that for all $\mathbf{i}$, $\mathbf{j}\in\mathbb{N}^N$ the joint probability distribution $\nu_{\mathbf{i},\mathbf{j}}$ of $X_{\mathbf{i}}$ and $X_{\mathbf{j}}$ satisfies

$$\exists\epsilon_1\in(0,1],\ \nu_{\mathbf{i},\mathbf{j}}\left(B(x,h_{\mathbf{n}})\times B(x,h_{\mathbf{n}})\right)=(p_{h_{\mathbf{n}}}^x)^{1+\epsilon_1},\tag{2.1}$$

where $p_{h_{\mathbf{n}}}^x=P(X\in B(x,h_{\mathbf{n}}))=\mu(B(x,h_{\mathbf{n}}))$. We recall that $\nu_{\mathbf{i},\mathbf{j}}$ is the joint distribution of $(X_{\mathbf{i}},X_{\mathbf{j}})$.

Remark 1. Note that this condition leads to $\nu_{\mathbf{i},\mathbf{j}}\left(B(x,h_{\mathbf{n}})\times B(x,h_{\mathbf{n}})\right)-\left(p_{h_{\mathbf{n}}}^x\right)^2=(p_{h_{\mathbf{n}}}^x)^2\left((p_{h_{\mathbf{n}}}^x)^{\epsilon_1-1}-1\right)\geq 0$ so $\left|\nu_{\mathbf{i},\mathbf{j}}\left(B(x,h_{\mathbf{n}})\times B(x,h_{\mathbf{n}})\right)-\left(p_{h_{\mathbf{n}}}^x\right)^2\right|=(p_{h_{\mathbf{n}}}^x)^2\left((p_{h_{\mathbf{n}}}^x)^{\epsilon_1-1}-1\right)\leq(p_{h_{\mathbf{n}}}^x)^{1+\epsilon_1}\leq 1$. Then, it can be link with the classical local dependence condition met in the literature when X and $(X_{\mathbf{i}},X_{\mathbf{j}})$ admitted respectively the densities f and $f_{\mathbf{i},\mathbf{j}}$:

$$\left|f_{\mathbf{i},\mathbf{j}}(x,y)-f(x)f(y)\right|\leq C,$$

for some constant $C>0$ and for all x, $y\in\mathcal{E}$ and $\mathbf{i}$, $\mathbf{j}\in\mathbb{N}^N$, $\mathbf{i}\neq\mathbf{j}$.

Actually, this condition can be replace by the more general version (which is not necessary here):

$$\exists\epsilon_1\in(0,1],\ P\left((X_{\mathbf{i}},X_{\mathbf{j}})\in B(x,h_{\mathbf{n}})\times B(y,h_{\mathbf{n}})\right)=(p_{h_{\mathbf{n}}}^x p_{h_{\mathbf{n}}}^y)^{\frac{1+\epsilon_1}{2}};$$

then, $\left|\nu_{\mathbf{i},\mathbf{j}}\left(B(x,h_{\mathbf{n}})\times B(y,h_{\mathbf{n}})\right)-p^x_{h_{\mathbf{n}}}\,p^y_{h_{\mathbf{n}}}\right|=\left(p^x_{h_{\mathbf{n}}}\,p^y_{h_{\mathbf{n}}}\right)\left(\left(p^x_{h_{\mathbf{n}}}\,p^y_{h_{\mathbf{n}}}\right)^{\frac{\epsilon_1-1}{2}}-1\right)\leq\left(p^x_{h_{\mathbf{n}}}p^y_{h_{\mathbf{n}}}\right)^{\frac{1+\epsilon_1}{2}}\leq 1$.

Such local dependency condition is necessary to reach the same rate of convergence as in the *i.i.d.* case.

2.2.2. Mixing conditions

Another complementary dependency condition concerned the *mixing condition* which measures the dependency by means of α-mixing. We assume that $\left(Z_{\mathbf{i}},\ \mathbf{i}\in\mathbb{Z}^N\right)$ satisfies the following mixing condition: there exists a function $\varphi(t)\downarrow 0$ as $t\to\infty$, such that for $E,\ E^{'}$ subsets of $\mathbb{Z}^N$ with finite cardinals,

$$\alpha\left(\mathcal{B}(E),\mathcal{B}\left(E^{'}\right)\right)=\sup_{B\in\mathcal{B}(E),\,C\in\mathcal{B}\left(E^{'}\right)}\left|\mathbf{P}(B\cap C)-\mathbf{P}(B)\,\mathbf{P}(C)\right|$$

$$\leq\chi\left(\mathrm{Card}\,(E)\,,\mathrm{Card}\left(E^{'}\right)\right)\psi\left(\mathrm{dist}\left(E,E^{'}\right)\right),\tag{2.2}$$

where $\mathcal{B}(E)$(*resp.* $\mathcal{B}\left(E^{'}\right)$) denotes the Borel σ-field generated by $(Z_{\mathbf{i}},\ \mathbf{i}\in E)$ (*resp.* $\left(Z_{\mathbf{i}},\ \mathbf{i}\in E^{'}\right)$), Card$(E)$ (*resp.* Card$\left(E^{'}\right)$) the cardinality of E (*resp.* $E^{'}$), dist$\left(E,\ E^{'}\right)$ the Euclidean distance between E and $E^{'}$ and $\chi:\mathbb{N}^2\to\mathbb{R}^+$ is a nondecreasing symmetric positive function in each variable. Throughout the paper, it will be assumed that χ satisfies either

$$\chi(n,m)\leq C\min(n,m)\,,\quad\forall n,m\in\mathbb{N}\tag{2.3}$$

or

$$\chi(n,m)\leq C\,(n+m+1)^{\widetilde{\beta}}\,,\qquad\forall n,m\in\mathbb{N}\tag{2.4}$$

for some $\widetilde{\beta}\geq 1$ and some $C>0$. If $\chi\equiv 1$, then $Z_{\mathbf{i}}$ is called strongly mixing. Many stochastic processes, among them various useful time series models satisfy strong mixing properties, which are relatively easy to check . Conditions (2.3)-(2.4) are weaker than strong mixing condition and have been used for finite dimensional variables in (for example) Tran [33], Carbon et al. [6, 7] and Biau and Cadre [3]. We refer to Doukhan [16] and Rio [31] for discussion on mixing and examples.

Concerning the function $\psi(.)$, we will only study the case where $\psi(i)$ tends to zero at a polynomial rate, ie.

$$\psi(i)\leq Ci^{-\theta},\text{ for some }\theta>0.\tag{2.5}$$

In the following, we denote by:

$$\theta_1=\frac{-\theta+N}{N(1+2\widetilde{\beta})-\theta},\ \theta_2=\frac{-\theta+2N}{2N(\beta+1)-\theta},$$

$$\theta_3 = \frac{-\theta + N}{N(1 + 2\beta + 2\widetilde{\beta}) - \theta}.$$

$$\theta_1^* = \frac{2N - \theta}{4N - \theta}, \ \theta_2^* = \frac{-\theta + N}{N(2\widetilde{\beta} + 3) - \theta},$$

$$\theta_3^* = \frac{-\theta + 2N}{-\theta + 2N(\beta + 2)}, \ \theta_4^* = \frac{-\theta + N}{N(3 + 2\beta + 2\widetilde{\beta}) - \theta}.$$

Remark 2. The results obtained below can be extend to the case where $\psi(i)$ tends to zero at an exponential rate: i.e $\psi(i) = C\ \exp(-si)$ for some $s > 0$.

- Each of the two dependence measures have the following specificity: if the first one control the local dependence, the second one control the dependence of sites which are far from each other.

- Clearly, one has, for a fixed $h_{\mathbf{n}}$ (note that the same argument can be easily generalized to the case where one deals with two different bandwidths) :

  $$\alpha(\|\mathbf{i} - \mathbf{j}\|) \geq \|g_{\mathbf{i},\mathbf{j}}\|_\infty$$

  with

  $$g_{\mathbf{i},\mathbf{j}}(x, y) = \nu_{\mathbf{i},\mathbf{j}}\left(B(x, h_{\mathbf{n}}) \times B(y, h_{\mathbf{n}})\right) - p_{h_{\mathbf{n}}}^x\, p_{h_{\mathbf{n}}}^y.$$

2.3. Assumptions on the kernel

We assume that the kernel $K : \mathbb{R} \to \mathbb{R}^+$ is of integral 1 and is such that:

HK_1 : there exist two constants $0 < C_1 < C_2 < \infty$:

$$C_1 I_{[0,1]} \leq K \leq C_2 I_{[0,1]}.$$

where $I_{[0,1]}$ is the indicator function in $[0, 1]$.
or

The support of K is $[0, 1]$, the derivative K' of K exists and satisfies

$$-\infty < C_1 \leq K' \leq C_2 < 0 \text{ and } -\exists C > 0, \exists \varepsilon_0 > 0, \forall \varepsilon < \varepsilon_0, \int_0^\varepsilon \mu(B(x, z))\, dz > C\varepsilon\mu(B(x, \varepsilon)).$$

In some cases, we will assume that:

HK_2 : K is a Lipschitz function.

3. Main Results

This section is devoted to the study of the consistency of the regression function: first, locally at a given point x of $\mathcal{E}$ and secondly, uniformly in the set $\mathcal{C}$.

3.1. Local convergence of the regression function

We study here the consistency of the regression function r at a given $x \in \mathcal{E}$. In this intention, we will use the assumption:

HF_1 - The regression function r is continuous at $x \in \mathcal{E}$,

and the following preliminary result (proved in the Appendix):

Lemma 3. *Let $G_{\mathbf{n}} = f_{\mathbf{n}}$ or $\varphi_{\mathbf{n}}$. If assumptions (2.1), (2.2) and HK_1 or HK_2 hold, then:*

$$\lim_{\mathbf{n}\to\infty} (\widehat{\mathbf{n}} p^x_{h_{\mathbf{n}}}) var(G_{\mathbf{n}}(x)) < \infty, \;\; x \in \mathcal{E}$$

as soon as the condition (2.3) or (2.4) is satisfied with $\sum_{i=1}^{\infty} i^{N-1}(\varphi(i))^a < \infty$, for some $0 < a < 1/2$.

As consequences of Lemma 3 we have the two following theorems.

Theorem 4. *Under the assumptions HF_1, HK_1, (2.1), (2.2), (2.5), $\widehat{\mathbf{n}} p^x_{h_{\mathbf{n}}} / \log \widehat{\mathbf{n}} \to \infty$ and if the mixing satisfies:*

- *conditions (2.3) and $\theta > 2N$*

or

- *conditions (2.4), $\theta > N(1 + 2\widetilde{\beta})$ and $\left(\widehat{\mathbf{n}}(p^x_{h_{\mathbf{n}}} / \log \widehat{\mathbf{n}})^{\theta_1}\right) \to \infty$*

then,

$$|r_{\mathbf{n}}(x) - r(x)| \text{ converges in probability to } 0. \tag{3.1}$$

The next results give the strong convergence of $r_{\mathbf{n}}$ under additional conditions. We set $g(\mathbf{n}) = \prod_{i=1}^{N} (\log n_i)(\log \log n_i)^{1+\epsilon}$, then we have $\sum_{\mathbf{n}\in\mathbb{N}^N} 1/\left(\widehat{\mathbf{n}}\, g(\mathbf{n})\right) < \infty$.

Theorem 5. *Under the assumptions HF_1, HK_1, (2.1), (2.2), (2.5), $\widehat{\mathbf{n}} p^x_{h_{\mathbf{n}}} / \log \widehat{\mathbf{n}} \to \infty$ and if the mixing verifies:*

- *the conditions (2.3), $\theta > 4N$ and* $\left(\widehat{\mathbf{n}} \left(\frac{p^x_{h_{\mathbf{n}}}}{\log \widehat{\mathbf{n}}}\right)^{\theta_1^*} g(\mathbf{n})^{\frac{2N}{4N-\theta}}\right)^{\frac{4N-\theta}{2N}} \to \infty$

  or

- *the conditions (2.4), $\theta > N(3 + 2\widetilde{\beta})$ and* $\left(\widehat{\mathbf{n}} \left(\frac{p^x_{h_{\mathbf{n}}}}{\log \widehat{\mathbf{n}}}\right)^{\theta_2^*} g(\mathbf{n})^{\frac{2N}{N(2\widetilde{\beta}+3)-\theta}}\right)^{\frac{N(2\widetilde{\beta}+3)-\theta}{2N}} \to \infty$

then,

$$|r_{\mathbf{n}}(x) - r(x)| \text{ converges almost surely to } 0. \tag{3.2}$$

3.2. Uniform convergence of the estimator over a set.

We consider a set $\mathcal{C}$ such that $\mathcal{C} \subset \mathcal{C}_{\mathbf{n}}$ where $\mathcal{C}_{\mathbf{n}} = \bigcup_{k=1}^{d_{\mathbf{n}}} B(t_k, \rho_{\mathbf{n}})$ (note that such set $\mathcal{C}_{\mathbf{n}}$ can always be built), $d_{\mathbf{n}} > 0$ is some integer, $t_k \in \mathcal{E}$, $k = 1, ..., d_{\mathbf{n}}$, and $\rho_{\mathbf{n}} > 0$. We assume that $\mathcal{C}$ is such that:

H_1 - $\sup_{x\in\mathcal{C}} p^x_{h_{\mathbf{n}}} = \Gamma(h_{\mathbf{n}}) > 0$ exists

H_2 - $d_{\mathbf{n}} = \widehat{\mathbf{n}}^{\beta}$ and $\rho_{\mathbf{n}} \leq (h_{\mathbf{n}})^{\kappa} \left(\log \widehat{\mathbf{n}}/(\widehat{\mathbf{n}}\Gamma(h_{\mathbf{n}}))\right)^{1/2}$, $\beta > 0 \kappa > 1$.

HF_2 - The regression function r is uniformly continuous on $\mathcal{C}$.

Under the dependence conditions (2.1), (2.2) and (2.5), we get the following weak and strong uniform consistency results:

Lemma 6. *Let $G_{\mathbf{n}} = f_{\mathbf{n}}$ or $\varphi_{\mathbf{n}}$. Under the condition of Lemma 3 and if H_1 is satisfied, we get*

$$\lim_{\mathbf{n}\to\infty} (\widehat{\mathbf{n}}\Gamma(h_{\mathbf{n}}))\mathbf{var}(G_{\mathbf{n}}(x)) < \infty, \;\; x \in \mathcal{C}.$$

The two following theorems are consequences of Lemma 6.

Theorem 7. *Under conditions HK_1, HF_2, and if $\widehat{\mathbf{n}}\Gamma(h_{\mathbf{n}})/\log \widehat{\mathbf{n}} \to \infty$, H_1, H_2 and*

- *the conditions (2.3), $\theta > 2N(\beta + 1)$ and $\left(\widehat{\mathbf{n}}(\Gamma(h_{\mathbf{n}})/\log \widehat{\mathbf{n}})^{\theta_2}\right) \to \infty$*

  or

- *the conditions (2.4), $\theta > N(1 + 2\beta + 2\widetilde{\beta})$ and $\left(\widehat{\mathbf{n}}(\Gamma(h_{\mathbf{n}})/\log \widehat{\mathbf{n}})^{\theta_3}\right) \to \infty$*

hold, we have

$$\sup_{x\in\mathcal{C}} |r_{\mathbf{n}}(x) - r(x)| \textit{ converges in probability to } 0. \tag{3.3}$$

Theorem 8. *Under conditions HK_1, HF_2 and if $\widehat{\mathbf{n}}\Gamma(h_{\mathbf{n}})/\log \widehat{\mathbf{n}} \to \infty$, H_1, H_2 and*

- *the conditions (2.3), $\theta > 2N(\beta + 2)$ and $\left(\widehat{\mathbf{n}}(\Gamma(h_{\mathbf{n}})/\log \widehat{\mathbf{n}})^{\theta_3^*} (g(\mathbf{n}))^{\frac{2N}{-\theta+2N(\beta+2)}}\right) \to \infty$*

  or

- *the conditions (2.4), $\theta > N(3+2\beta+2\widetilde{\beta})$ and $\left(\widehat{\mathbf{n}}(\Gamma(h_{\mathbf{n}})/\log \widehat{\mathbf{n}})^{\theta_4^*} (g(\mathbf{n}))^{\frac{2N}{-\theta+N(2\beta+2\widetilde{\beta}+3)}}\right) \to \infty$*

hold, we have

$$\sup_{x\in\mathcal{C}} |r_{\mathbf{n}}(x) - r(x)| \text{ converges almost surely to } 0. \tag{3.4}$$

The rate of convergence of $r_{\mathbf{n}}$ is given in the next theorems under the following additional conditions on the model.

HF_3 - We assume in the following theorem that r is a Lipschitz function.

Theorem 9. *Under the conditions of Theorem 7 except that HK_1 and HF_2 are respectively replaced by HK_2 and HF_3, we have*

$$\sup_{x\in\mathcal{C}} |r_{\mathbf{n}}(x) - r(x)| = \mathcal{O}\left(h_{\mathbf{n}} + \sqrt{\frac{\log \widehat{\mathbf{n}}}{\Gamma(h_{\mathbf{n}})\widehat{\mathbf{n}}}}\right) \text{ in probability.} \tag{3.5}$$

The strong rate of convergence of $r_{\mathbf{n}}$ follows in the two cases of mixing.

Theorem 10. *Under the conditions of Theorem 8 except that HK_1 and HF_2 are respectively replaced by HK_2 and HF_3, we have:*

$$\sup_{x\in\mathcal{C}} |r_{\mathbf{n}}(x) - r(x)| = \mathcal{O}\left(h_{\mathbf{n}} + \sqrt{\frac{\log \widehat{\mathbf{n}}}{\Gamma(h_{\mathbf{n}})\widehat{\mathbf{n}}}}\right) \ a.s. \tag{3.6}$$

4. Simulations

As it was raised ealier in this paper, if our estimator looks like its *i.i.d.*'s counterpart, there is a slight difference in practice. Indeed, as mentioned in Dabo-Niang et al. [13], one must take into account the spatial dependency which is supposed here to be measured by mixing condition. In this Section, we are interested both by a way of applying our estimator illustred by some simulation tudies.

4.1. The mixing condition in practice.

We are dealing with a non parametric spatial dependence measure. We recall that for any couple of sites $(\mathbf{i},\mathbf{j})$

$$\alpha\left(||\mathbf{i}-\mathbf{j}||\right) \leq \chi(1,1)\,\psi\left(||\mathbf{i}-\mathbf{j}||\right).$$

For the sake of simplicity, let us consider the strong mixing case (which corresponds to the case where $\chi \equiv 1$). Then, we have assumed that $\psi\left(||\mathbf{i}-\mathbf{j}||\right)$ tend to zero at a polynomial rate (or respectively at exponential rate). That is, we can say that $\psi\left(||\mathbf{i}-\mathbf{j}||\right) \leq C||\mathbf{i}-\mathbf{j}||^{-\theta}$ (respectively $\psi\left(||\mathbf{i}-\mathbf{j}||\right) \leq \exp(-\theta||\mathbf{i}-\mathbf{j}||)$) for some $\theta > 0$. Thus, this assumption, might be taken into account when using the regression estimator given in Section 2. This argument leads us to say that actually, we are dealing with the following regression estimator.

4.1.1. The spatial regression estimator in practice.

For all $x_{\mathbf{j}}$ (which could be observed a site $\mathbf{j}$),

$$r_{\mathbf{n}}(x_{\mathbf{j}}) = \begin{cases} \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i}} W_{\mathbf{i},x_{\mathbf{j}}} & \text{if } \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} W_{\mathbf{i},x_{\mathbf{j}}} \neq 0; \\ \frac{1}{\hat{\mathbf{n}}} \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i}}\, \mathbf{I}_{V_{\mathbf{j}}}(\mathbf{i}) & \text{else.} \end{cases}$$

where $\mathbf{I}_{V_{\mathbf{j}}}$ is the indicator function of the set $V_{\mathbf{j}} = \left\{\varphi\left(||\mathbf{i}-\mathbf{j}||\right) > C\,||\mathbf{i}-\mathbf{j}||^{-\theta}\right\}$ and

$$W_{\mathbf{i},x_{\mathbf{j}}} = \frac{K\left(d(X_{\mathbf{i}},x_{\mathbf{j}})h_{\mathbf{n}}^{-1}\right)\mathbf{I}_{V_{\mathbf{j}}}(\mathbf{i})}{\sum_{\mathbf{m}\in\mathcal{I}_{\mathbf{n}}} K\left(d(X_{\mathbf{m}},x_{\mathbf{j}})h_{\mathbf{n}}^{-1}\right)\mathbf{I}_{V_{\mathbf{j}}}(\mathbf{m})}.$$

So we have:

$$r_{\mathbf{n}}(x_{\mathbf{j}}) = \begin{cases} \frac{\sum_{\mathbf{i}\in V_{\mathbf{j}}} Y_{\mathbf{i}}\, K\left(d(X_{\mathbf{i}},x_{\mathbf{j}})h_{\mathbf{n}}^{-1}\right)}{\sum_{\mathbf{i}\in V_{\mathbf{j}}} K\left(d(X_{\mathbf{i}},x_{\mathbf{j}})h_{\mathbf{n}}^{-1}\right)} & \text{if } \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} W_{\mathbf{i},x_{\mathbf{j}}} \neq 0; \\ \frac{1}{\hat{\mathbf{n}}} \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i}}\, \mathbf{I}_{V_{\mathbf{j}}}(\mathbf{i}) & \text{else.} \end{cases}$$

Note that $V_{\mathbf{j}}$ is the set of Card($V_{\mathbf{j}}$) nearest neighbors of $\mathbf{j}$.

Remark 11.

1. The choice of the bandwidth (even in finite or infinite dimensional setting) is a crucial question in non-parametric estimation. We propose to chose the optimal bandwidth by using cross-validation procedure.
2. Another interesting problem is the estimation of the sets $V_{\mathbf{j}}$'s. This problem is the subject of another work in progress. Nevertheless, for simplicity, we will consider that $V_{\mathbf{j}}$'s is a set of an arbitrary $k_{\mathbf{n}}$ number of nearest neighbors (by meaning of the euclidean distance).

Let $k_{\mathbf{n}}$ be an integer. Then, the regression function estimation at point $x_{\mathbf{j}}$ is obtained by using the following algorithm:

4.1.2. Algorithm for spatial regression estimation based on nearest neighbors.

1. Compute the optimal bandwith, $h_{k_{\mathbf{n}},opt}$, by using cross-validation procedure.
2. Take the $k_{\mathbf{n}}$ nearest neighboors of each site.
3. Compute respectively the of $k_{\mathbf{n}}$'s reals $K\left(\frac{d(X_{\mathbf{j}},X_{\mathbf{i}})}{h_{k_{\mathbf{n}},opt}}\right)$ and $Y_{\mathbf{i}}K\left(\frac{d(X_{\mathbf{j}},X_{\mathbf{i}})}{h_{k_{\mathbf{n}},opt}}\right)$, $\mathbf{i} \in V_{\mathbf{j}}$: $\sum_{\mathbf{i}\in V_{\mathbf{j}}} K\left(\frac{d(X_{\mathbf{j}},X_{\mathbf{i}})}{h_{k_{\mathbf{n}},opt}}\right)$, $\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i}} W_{\mathbf{i},x_{\mathbf{j}}}$.
4. Compute $r_{\mathbf{n}}(x_{\mathbf{j}})$.

This algorithm is illustrated in the following simulation studies. In the following, we will denote by **Piid**, the procedure of estimation of Ferraty and Vieu [18] and Dabo-Niang and Rhomari [12] and by **Psdep,** our procedure of estimation. Note that **Psdep** and **Piid**, coincide as soon as $k_{\mathbf{n}} = \hat{\mathbf{n}}$.

4.2. Simulations studies

In order to illustrate our results, we have done some simulations based on observations $(X_{i,j}, Y_{i,j})$, $0 \le i,\, j \le 25$ such that $\forall\, i, j,$:

$$X_{i,j}(t) = A_{i,j} * (t - 0.5)^2 + B_{i,j}$$

and

$$Y_{i,j} = 4A_{i,j}^2 + \varepsilon_{i,j}, \tag{4.1}$$

where $A = (A_{i,j})$, $B = (B_{i,j})$ and $\varepsilon = (\varepsilon_{i,j})$ are random variables which will be specified later on. Note that here we have $r(X) = 4.X''$ (where f'' denotes the second derivatives of a funtion f)). We are first (on Section 1.2.1.) interesting with the estimation of Model (4.1) based on *i.i.d.* observations $Z_{\mathbf{i}} = (X_{\mathbf{i}}, Y_{\mathbf{i}})$ (the sequences A, B and ε are then i.i.d. random variables); after that we deal on Section 1.2.2, with Model (4.1) generated with the spatial dependence structure.

In each case, we have done 30 simulations of Model (4.1) and compared the quality of estimation of **Piid** and **Psdep**. The quality of estimation is meaning by coefficient of determination R^2. The results are presented in Table 1 where each figure includes on one hand, 30 points (on red) representing the 30 values of the R^2 obtained by **Piid**. And on the other hand, the 30 curves defined by $((k_{\mathbf{n}}, R^2(k_{\mathbf{n}}))$ discretized in points $k_{\mathbf{n}} = 8 + 5.\ell$, $\ell = 1, 2, ..., 29$ obtained by **Psdep**.

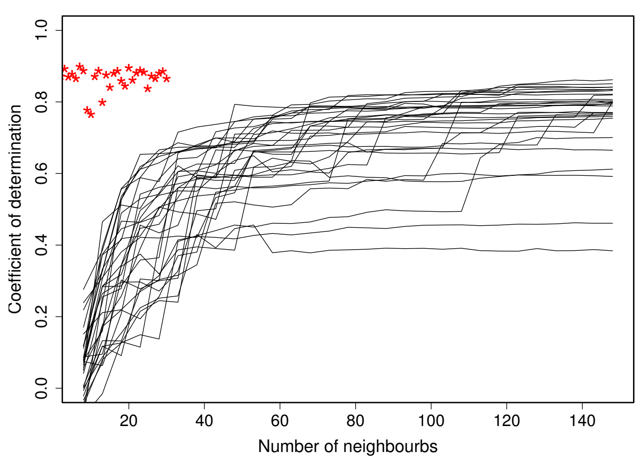
4.2.1. Model (4.1) with i.i.d. observations

In this Section, Model (4.1) is simulated with *i.i.d.* observations. Namely, the sequences $(A_{i,j})$, $(B_{i,j})$ and $(\varepsilon_{i,i})$ are $25{\times}25$ *i.i.d.* random variables such that $\forall i, j$, $A_{i,j} \sim \mathcal{N}(0, 1)$, $B_{i,j} \sim \mathcal{N}(0, .1)$ and $\varepsilon_{i,j} \sim \mathcal{N}(0, 2)$. We have done 30 simulations of this model, the results are presented in Table 1, Figure A.

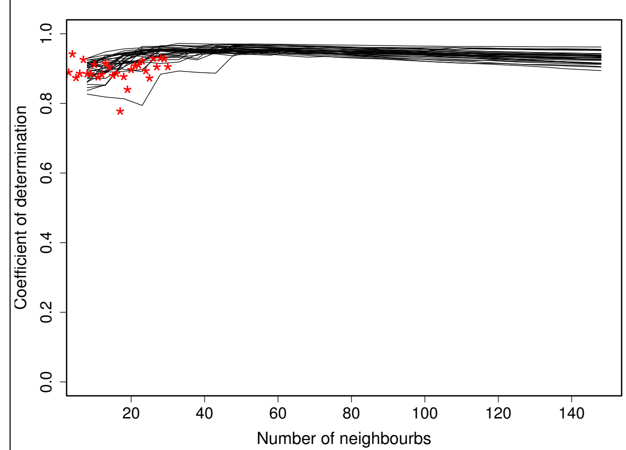
These results shows that (as it is expected), procedure **Piid** leads to better estimation of Model (4.1) than **Psdep**. Futhermore, the quality of the estimation obtained by **Psdep** is improved as $k_{\mathbf{n}}$ increases and tends to the quality of estimation of **Piid**. That is explained by the fact that as $k_{\mathbf{n}}$ increases, one tends to the situation where **Psdep** and **Piid** coincide ($k_{\mathbf{n}} = \hat{\mathbf{n}}$).

4.2.2. With spatial dependency

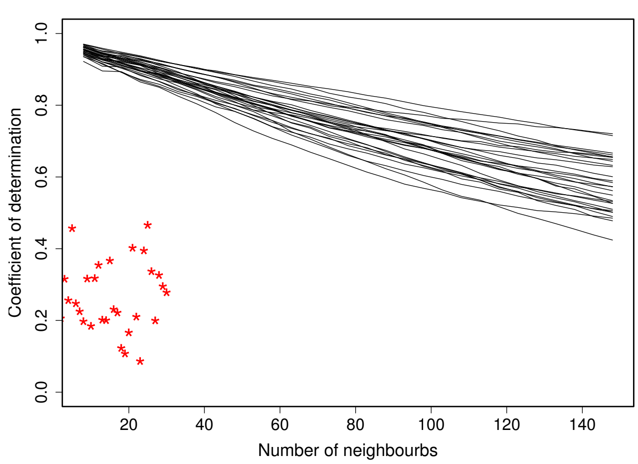
This time, Model (4.1) is simulated with spatial dependence structure. Thereafter, we denote by $GRF(m, \sigma^2, s)$ a stationary Gaussian random field with mean m and covariance function defined by $C(h) = \sigma^2 \exp(-(\frac{\|h\|}{s})^2)$, $h \in \mathbb{R}^2$ and $s > 0$. Then, we have then simulated Model (4.1) with $A = D * \sin(\frac{G}{2} + .5)$, $B = GRF(2.5, 5, 3)$, $\varepsilon = GRF(0, .1, 5)$,



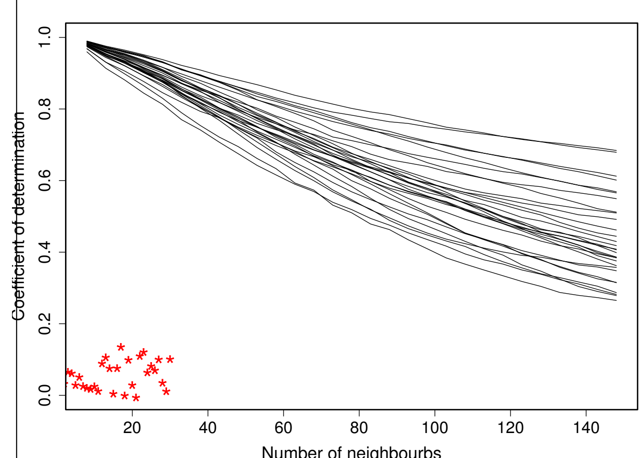
A. With i.i.d. observations



B. With spatial dependency, with a=50



C. With spatial dependency,with a=20



D. With spatial dependency,with a=5

Table 1: Values of the coefficient of determination R^2

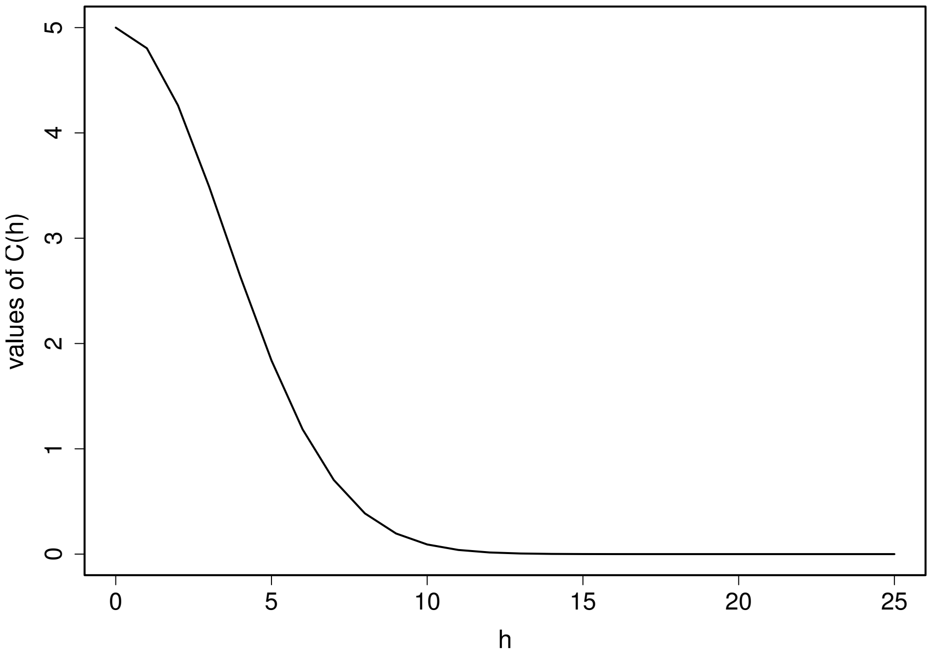


Figure 4.1: Covariance function with $\sigma^2 = 5$ and $s = 5$

$G = GRF(0, 5, 3)$ and $D_{\mathbf{i}} = \frac{1}{\hat{\mathbf{n}}} \sum_{\mathbf{j}} \exp\left(-\frac{\|\mathbf{i}-\mathbf{j}\|}{a}\right)$
($D_{(i,j)} = \frac{1}{25\times 25} \sum_{1\leq m,t \leq 25} \exp\left(-\frac{\|(i,j)-(m,t)\|}{a}\right)$). The function D is here to ensure and control the spatial mixing condition (even if using the Gaussian Random Fields also brings some spatial dependency). Indeed, our model can be seen verifying a mixing condition with $\psi(h) \to 0$ at exponential rate. Then, the greater is a, the weaker is the spatial dependency. Futhermore, if $a \to \infty$, $D_{\mathbf{i}} \to 1$.

Now, let us respectively consider cases $a = 50$, 20, 5. The case $a = 50$ corresponds to the one we discuss just before since $D_{\mathbf{i}} \simeq 1$. The results are presented on Table 1-Figure B where, whatever the values of $k_{\mathbf{n}}$, one has a good quality of estimation both with **Psdep** and **Piid** and the values are almost equal. The fact that quality of estimation by **Piid** is as good (despite the existence of dependence) is explained by the high valued of a and the number of independent observations is then not negligeable. Actually, this later case corresponds to $A \simeq \sin(\frac{G}{2} + .5)$ and Model (4.1) is based both on spatial dependent observations and nearly *i.i.d.* observations. In fact, since (in these conditions) Model (4.1) is based on Gaussian random fields with covariance function C and scale $s \leq 5$ (see

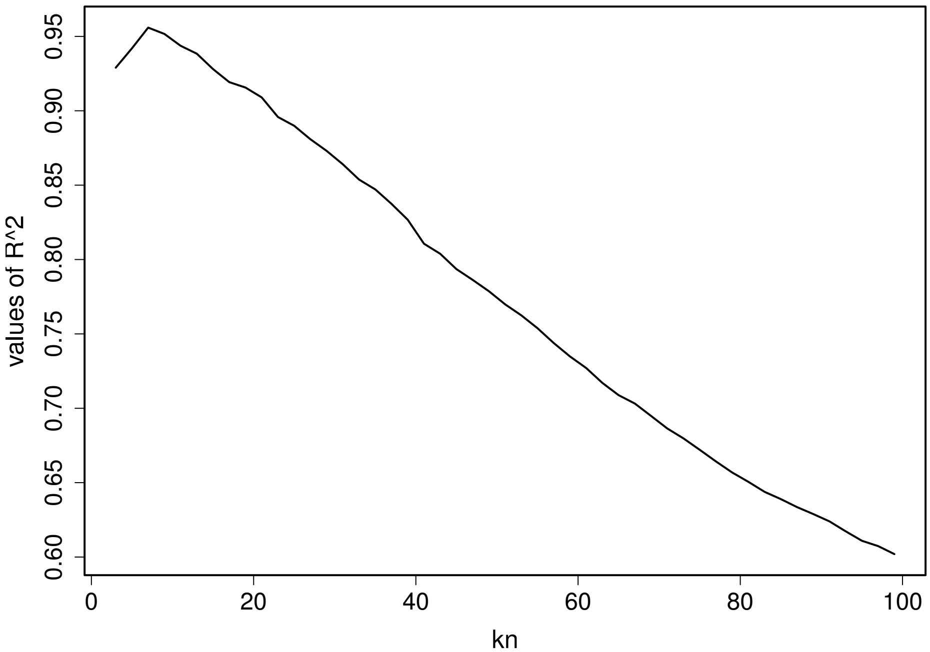
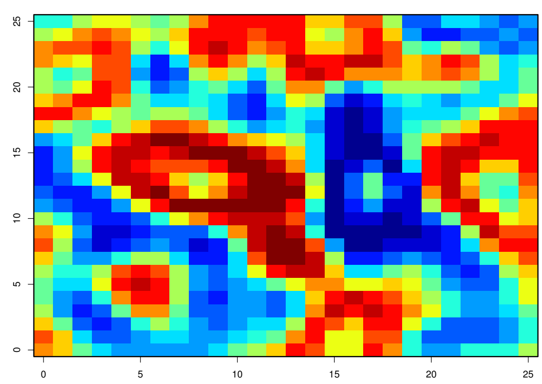
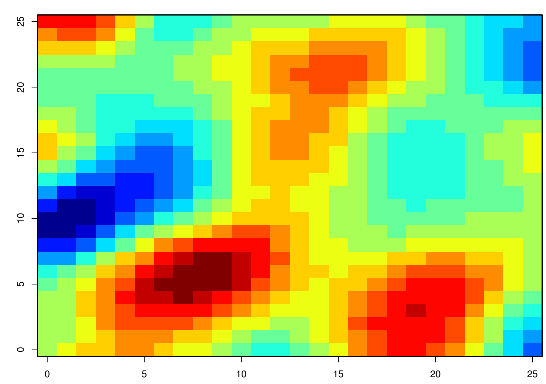


Figure 4.2: kn versus R^2 with $a = 20$

Figure 4.1), observations of sites $\mathbf{i}$ and $\mathbf{j}$ with $\|\mathbf{i}-\mathbf{j}\| < 15$ are spatial dependent and nearly independent from $\|\mathbf{i}-\mathbf{j}\| \geq 15$. So, our observations are a mixture of *i.i.d.* and dependent observations. Thus, to move away from independence, it suffises to lower the value of a. That is done in the context of Table1, figures C and D respectively with $a = 20$ and $a = 5$. One can see that the quality of estimation of **Piid** deteriorate as a decreases and is very bad with $a = 5$.

Other interesting results are the evolution of quality estimation of **Psdep** in Table 1, figures B, C and D which are different from Figure 4.1.A. In fact, as one can see on Figure 4.2 there is an optimal $k_{\mathbf{n}}$ around which the quality of estimation is better and quality is increasingly bad when away from this values and tends to that of **Piid**. These results are not visible in the previous figures the discretization is to coarse.





A realization of the random field A	A realiszation of the ramdom field B

Figure 4.3: Simulation

5. Conclusion and discussion

In this paper, we have developped a new method in non-parametric spatial modelling (for functional random fields). Then, when the observations are high-dimensional spatial data (as are curves), this method appears as a good alternative to existing ones.

More precisely, we have studied theoretically the asymptotic behavior of our method and, illustrate its pratical use through some finite size simulations. All this makes the proposal very attractive.

In addition, this work offers very interesting perspectives of investigation. In fact, as mentioned above, we have two main problems with our procedure of estimation: the choice of both the bandwith and $k_{\mathbf{n}}$. To solve this problem, we have chosen an optimal bandwith (using cross validation) for each fixed $k_{\mathbf{n}}$ as shows Figure 5.1 (and of course tends to the optimal bandwidth of **Piid** as $k_{\mathbf{n}}$ tends to $\hat{\mathbf{n}}$). Then, a question raised : " Does the results fundamentally change when choosing simultaneously $k_{\mathbf{n}}$ and the optimal bandwith?" So, an outlook of this work is the statement of theoretical properties with respect to the choice of these two parameters using cross-validation method in functional random fields modelling (as it is in the *i.i.d.* setting by for exampleRachdi and Vieu [29] for bandwidth selection).

Finally, this work is a step towards functional random fields models taking into account both the functional and spatial dependency feature of the data. The results obtained here are encouraging to pursue investigations in this topic. Namelly, in a work in progress, we aim to apply this method to apply spatial prediction and real data problem.

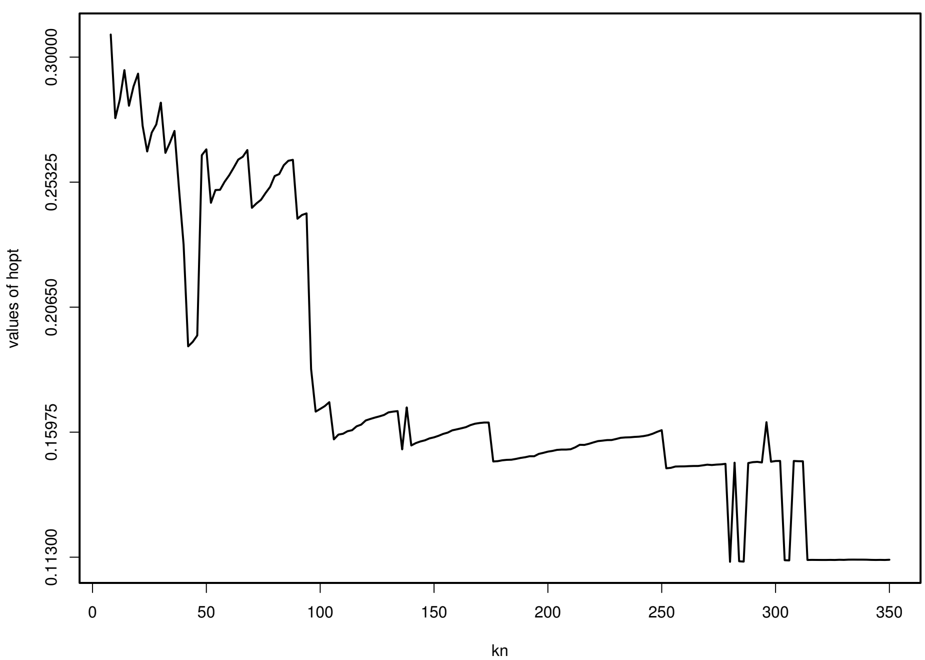


Figure 5.1: $k_{\mathbf{n}}$ versus $h_{k_n,opt}$ with $a = 20$: the value of hopt for **Piid** is 0.113

6. Appendix

This section is devoted to prove the consistency result stated in the previous sections. For that, we recall three lemmas which can be find to Carbon et al. [5] which will be used in the following. As previously, along this section we will denote by C a positive generic constant.

Lemma 12. *Suppose $E_1, ..., E_r$ be sets containing m sites each with $dist(E_i,\ E_j) \geq \gamma$ for all $i \neq j$ where $1 \leq i \leq r$ and $1 \leq j \leq r$. Suppose $Z_1, ..., Z_r$ is a sequence of real-valued r.v.'s measurable with respect to $\mathcal{B}(E_1), ..., \mathcal{B}(E_r)$ respectively, and Z_i takes values in $[a, b]$. Then there exists a sequence of independent r.v.'s $Z_1^*, ..., Z_r^*$ independent of $Z_1, ..., Z_r$ such that Z_i^* has the same distribution as Z_i and satisfies*

$$\sum_{i=1}^{r} E|Z_i - Z_i^*| \leq 2r(b-a)\chi((r-1)m, m)\psi(\gamma) \tag{6.1}$$

Lemma 13.
(i) Suppose that (2.2) holds. Denote by $\mathcal{L}_r(\mathcal{F})$ the class of $\mathcal{F}-$measurable r.v.'s X satisfying $\|X\|_r = (E|X|^r)^{1/r} < \infty$. Suppose $X \in \mathcal{L}_r(\mathcal{B}(E))$ and $X \in \mathcal{L}_r(\mathcal{B}(E'))$. Assume also that $1 \leq r,\ s,\ t < \infty$ and $r^{-1} + s^{-1} + t^{-1} = 1$. Then

$$|EXY - EXEY| \leq C\|X\|_r\|Y\|_s\{\chi(Card(E), Card(E'))\psi(dist(E, E'))\}^{1/t}. \tag{6.2}$$

(ii) For r.v.'s bounded with probability 1, the right-hand side of (6.2) can be replaced by $C\chi(Card(E), Card(E'))\psi(dist(E, E'))$.

Lemma 14.
If (2.5) holds for $\theta > 2N$, then

$$\sum_{i=1}^{\infty} i^{N-1}(\psi(i))^a < \infty \tag{6.3}$$

for some $0 < a < 1/2$.

6.1. Proofs of the results for the local convergence case.

Proof. of Lemma 3: Let

$$Z_{\mathbf{i},x} = \frac{K_{\mathbf{i}}}{\Delta_x} K\left(d\left(X_{\mathbf{i}}, x\right) h_{\mathbf{n}}^{-1}\right) - E\left(\frac{K_{\mathbf{i}}}{\Delta_x} K\left(d\left(X_{\mathbf{i}}, x\right) h_{\mathbf{n}}^{-1}\right)\right), \quad \mathbf{S_n} = \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} Z_{\mathbf{i},x},$$

where $\Delta_x = E\left(K\left(d\left(X,x\right)h_{\mathbf{n}}^{-1}\right)\right)$, $K_{\mathbf{i}} = 1$ (for the case of $f_{\mathbf{n}}$) or $K_{\mathbf{i}} = Y_{\mathbf{i}}$ (for the case of $\varphi_{\mathbf{n}}$).

We have $var(G_{\mathbf{n}}(x)) = var(\mathbf{S_n}/\widehat{\mathbf{n}})$, then

$$var(G_{\mathbf{n}}(x)) \leq \widehat{\mathbf{n}}^{-2}\left(\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} E(Z_{\mathbf{i},x}^2) + \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}\ \sum_{\mathbf{j}\in\mathcal{I}_{\mathbf{n}},\, i_k\neq j_k \text{ for some } k} |E(Z_{\mathbf{i},x}Z_{\mathbf{j},x})|\right).$$

Let us first consider the case where $K_{\mathbf{i}} = 1$ and set $J_{1,x,\mathbf{n}} = \left(\widehat{\mathbf{n}}p_{h_{\mathbf{n}}}^x\right)\left(\widehat{\mathbf{n}}^{-2}\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} E(Z_{\mathbf{i},x}^2)\right)$. Then, we have the inequality

$$J_{1,x,\mathbf{n}} \leq \frac{p_{h_{\mathbf{n}}}^x}{(\Delta_x)^2}\left[E(K^2\left(d\left(X,x\right)h_{\mathbf{n}}^{-1}\right))\right]$$

which leads by assumption HK_1 to:

$$J_{1,x,\mathbf{n}} \leq C\frac{p_{h_{\mathbf{n}}}^x}{\Delta_x}\frac{E\left(K\left(d\left(X,x\right)h_{\mathbf{n}}^{-1}\right)\right)}{\Delta_x} = C\frac{p_{h_{\mathbf{n}}}^x}{\Delta_x}.$$

Furthermore, it is easy to see that $0 < C \leq p_{h_{\mathbf{n}}}^x/\Delta_x \leq C'$ (see for example Lemma 4.3 or 4.4 of Ferraty and Vieu [18]) where C and C' are two real constants, so:

$$\lim_{\mathbf{n}\to\infty} J_{1,x,\mathbf{n}} < \infty.$$

To continue the proof, we set:

$$J_{2,x,\mathbf{n}} = \left(\widehat{\mathbf{n}}p_{h_{\mathbf{n}}}^x)\right)\left(\widehat{\mathbf{n}}^{-2}\left(\sum\sum_{0<dist(\mathbf{i},\mathbf{j})\leq m_{\mathbf{n}}} E(Z_{\mathbf{i},x}Z_{\mathbf{j},x}) + \sum\sum_{dist(\mathbf{i},\mathbf{j})>m_{\mathbf{n}}} E(Z_{\mathbf{i},x}Z_{\mathbf{j},x})\right)\right) = J_{2,x,\mathbf{n}}^1 + J_{2,x,\mathbf{n}}^2.$$

where $m_{\mathbf{n}}$ is a positive real (depending on $\mathbf{n}$).

Now, since

$$E(Z_{\mathbf{i},x}Z_{\mathbf{j},x}) \leq \frac{C}{(\Delta_x)^2}\left|P\left((X_{\mathbf{i}},X_{\mathbf{j}}) \in B(x,h_{\mathbf{n}}) \times B(x,h_{\mathbf{n}})\right)\right|,$$

by assumptions HK_1, where $0 < \epsilon_1 < 1$ one get:

$$E(Z_{\mathbf{i},x}Z_{\mathbf{j},x}) \leq \frac{C}{(\Delta_x)^2}(p_{h_{\mathbf{n}}}^x)^{1+\epsilon_1} \leq C(p_{h_{\mathbf{n}}}^x)^{\epsilon_1 - 1}$$

and by the way the inequality:

$$J_{2,x,\mathbf{n}}^1 \leq C(\widehat{\mathbf{n}}p_{h_{\mathbf{n}}}^x)\widehat{\mathbf{n}}m_{\mathbf{n}}^N\widehat{\mathbf{n}}^{-2}(p_{h_{\mathbf{n}}}^x)^{\epsilon_1-1} = Cm_{\mathbf{n}}^N {p_{h_{\mathbf{n}}}^x}^{\epsilon_1}.$$

Setting $m_{\mathbf{n}} = {p_{h_{\mathbf{n}}}^x}^{-(1-\gamma)\epsilon_1/\nu}$, where $\nu = -N - \epsilon_1 + (1-\gamma)\epsilon_1 N a^{-1}$ with γ and ϵ_1 some small positive numbers such that $a^{-1}\epsilon_1 - (N+\epsilon_1)(N(1-\gamma))^{-1} > 1$ (this is possible since $0 < a < 1/2$), we get:

$$J_{2,x,\mathbf{n}}^1 \leq C(p_{h_{\mathbf{n}}}^x)^{1-N(1-\gamma)/\nu},$$

thus $\lim_{\mathbf{n}\to\infty} J^1_{2,x,\mathbf{n}} = 0$ since $\nu > N(1-\gamma)$.

Let us turn to $J^2_{2,x,\mathbf{n}}$ and let $\gamma' = 1-(1-\gamma)\epsilon_1$, $\delta = 2(1-\gamma')/\gamma'$. Notice that $\gamma' = 2/(2+\delta)$ and $1-\gamma' = \delta/(2+\delta)$. We apply Lemma 2.1 of Tran [33]with $r = s = 2+\delta$; $h = (2+\delta)/\delta$ and get the inequality:

$$|E\left(Z_{\mathbf{i},x}Z_{\mathbf{j},x}\right)| \leq C\left(\frac{1}{(\Delta_x)^{2+\delta}}E^{2+\delta}\left[K\left(d\left(X_{\mathbf{k}},x\right)h_{\mathbf{n}}^{-1}\right)\right]\right)^{\gamma'}\left(\psi(\|\mathbf{i}-\mathbf{j}\|)\right)^{1-\gamma'};$$

which leads to:

$$J^2_{2,x,\mathbf{n}} \leq (\widehat{\mathbf{n}}^{-1}p^x_{h_\mathbf{n}})\sum\sum_{\|\mathbf{i}-\mathbf{j}\|>m_\mathbf{n}} C\left(\frac{1}{(\Delta_x)^{2+\delta}}E^{2+\delta}\left[K\left(d\left(X_{\mathbf{k}},x\right)h_{\mathbf{n}}^{-1}\right)\right]\right)^{\gamma'}\left(\psi(\|\mathbf{i}-\mathbf{j}\|)\right)^{1-\gamma'}.$$

$$\leq C(\widehat{\mathbf{n}}^{-1}p^x_{h_\mathbf{n}})(p^x_{h_\mathbf{n}})^{-\gamma'(1+\delta)}\sum\sum_{\|\mathbf{i}-\mathbf{j}\|>m_\mathbf{n}}\left(\psi(\|\mathbf{i}-\mathbf{j}\|)\right)^{1-\gamma'}$$

$$\leq C(\widehat{\mathbf{n}}^{-1}p^x_{h_\mathbf{n}})(p^x_{h_\mathbf{n}})^{-\gamma'(1+\delta)}\widehat{\mathbf{n}}\sum_{\|\mathbf{i}\|>m_\mathbf{n}}\left(\psi(\|\mathbf{i}\|)\right)^{1-\gamma'}$$

$$\leq C(p^x_{h_\mathbf{n}})^{-1+\gamma'}\sum_{\|\mathbf{i}\|>m_\mathbf{n}}\left(\varphi(\|\mathbf{i}\|)\right)^{1-\gamma'}.$$

It comes from assumption $\sum_{i=1}^{\infty} i^{N-1}(\psi(i))^a < \infty$ that $\psi(i) = o(i^{-N/a})$ and since φ is a decreasing function we have $\varphi(t) = o(t^{-N/a})$ as $t \to \infty$ and

$$\|\mathbf{i}\|^{\nu}\left(\psi(\|\mathbf{i}\|)\right)^{1-\gamma'} = \|\mathbf{i}\|^{\nu}o(\|\mathbf{i}\|^{-N(1-\gamma')/a}) = o(\|\mathbf{i}\|^{-N-\epsilon_1}),$$

because $\nu = -N-\epsilon_1+(1-\gamma)\epsilon_1 N a^{-1} > 0$. Then,

$$\sum_{\mathbf{i}\in\mathcal{I}_\mathbf{n}}\|\mathbf{i}\|^{\nu}\left(\psi(\|\mathbf{i}\|)\right)^{1-\gamma'} < \infty.$$

Furthermore $(p^x_{h_\mathbf{n}})^{1-\gamma'}m_n^{-\nu} = 1$ so:

$$\limsup J^2_{2,x,\mathbf{n}} \leq C\limsup\,(p^x_{h_\mathbf{n}})^{-1+\gamma'}\sum_{\|\mathbf{i}\|>m_\mathbf{n}}\left(\psi(\|\mathbf{i}\|)\right)^{1-\gamma'}$$

$$\leq C\limsup(p^x_{h_\mathbf{n}})^{-1+\gamma'}m_n^{-\nu}\sum_{\|\mathbf{i}\|>m_\mathbf{n}}\|\mathbf{i}\|^{\nu}\left(\psi(\|\mathbf{i}\|)\right)^{1-\gamma'}$$

$$\leq C\limsup\sum_{\|\mathbf{i}\|>m_\mathbf{n}}\|\mathbf{i}\|^{\nu}\left(\psi(\|\mathbf{i}\|)\right)^{1-\gamma'}.$$

This last term tends to zero as $m_\mathbf{n}$ tends to infinity. For the case $K_\mathbf{i} = Y_\mathbf{i}$ the proof is the same as the previous one since Y is bounded. This yields the proof. ■

Proof. of Theorem 4.
We consider the decomposition:

$$r_{\mathbf{n}}(x)-r(x)=\frac{1}{f_{\mathbf{n}}(x)}\left\{\varphi_{\mathbf{n}}(x)-E(\varphi_{\mathbf{n}}(x))-(r(x)-E(\varphi_{\mathbf{n}}(x)))\right\}-\frac{r(x)}{f_{\mathbf{n}}(x)}\left\{f_{\mathbf{n}}(x)-1\right\},\ x\in\mathcal{E}. \tag{6.4}$$

Let:

$$I_1(x)=\varphi_{\mathbf{n}}(x)-E(\varphi_{\mathbf{n}}(x)),$$

$$I_2(x)=r(x)-E(\varphi_{\mathbf{n}}(x)).$$

and

$$I_3(x)=f_{\mathbf{n}}(x)-1.$$

Then, we have for a given $\mathbf{k}\in\mathbb{N}^N$

$$\begin{aligned}I_2(x)=r(x)-E\left(\frac{Y}{\Delta_x}K\left(d\left(X,x\right)h_{\mathbf{n}}^{-1}\right)\right)&=r(x)-E\left(E\left(\frac{Y_{\mathbf{k}}}{\Delta_x}K\left(d\left(X_{\mathbf{k}},x\right)h_{\mathbf{n}}^{-1}\right)|X_{\mathbf{k}}\right)\right)\\&=r(x)-E\left(r(X_{\mathbf{k}})\left(\frac{K\left(d\left(X_{\mathbf{k}},x\right)h_{\mathbf{n}}^{-1}\right)}{\Delta_x}\right)\right)\\&=E\left((r(x)-r(X_{\mathbf{k}}))\left(\frac{K\left(d\left(X_{\mathbf{k}},x\right)h_{\mathbf{n}}^{-1}\right)}{\Delta_x}\right)\right).\end{aligned}$$

Since the support of the function K is $[0,1]$, we have:

$$r(x)-r(X_{\mathbf{k}})\leq\sup_{u\in B(x,h_{\mathbf{n}})}|r(x)-r(u)|.$$

So, $I_2(x)\leq\sup_{u\in B(x,h_{\mathbf{n}})}|r(x)-r(u)|$ converges to zero by the continuity assumption of r at x.

We now focus on the convergence of $I_1(x)$. Note that the proof of $I_3(x)$ is derive from the one of $I_1(x)$ by setting $Y_{\mathbf{i}}=1$.

Let :

$$Q_{\mathbf{n}}(x)=\varphi_{\mathbf{n}}\left(x\right)-E\left(\varphi_{\mathbf{n}}\left(x\right)\right)=\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}Z_{\mathbf{i},\,\mathbf{n},\,x}-E\left(Z_{\mathbf{i},\,\mathbf{n}}\right)=\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}Z'_{\mathbf{i},\,\mathbf{n},\,x},\qquad x\in\mathcal{E},$$

where

$$Z_{\mathbf{i},\,\mathbf{n},\,x}=\frac{Y_{\mathbf{i}}\Delta_{\mathbf{i}}}{\widehat{\mathbf{n}}},\ \Delta_{\mathbf{i}}=\frac{K\left(d\left(X_{\mathbf{i}},x\right)h_{\mathbf{n}}^{-1}\right)}{EK\left(d\left(X,x\right)h_{\mathbf{n}}^{-1}\right)}$$

and $p \geq 1$ and integer. Without loss of generality assume that for some integers $t_1, ..., t_N$,

$$n_i = 2pt_i, \; i = 1, ..., N. \tag{6.5}$$

The following spatial blocking idea here is that of Tran [33]. If it is not the case that $n_i = 2t_ip, \; i = 1, ..., N$ for some integers $t_1, ..., t_N$, then a term say $T(\mathbf{n}, x, 2^N + 1)$, containing all the $Z_{\mathbf{i},\mathbf{n},x}$'s at the ends not included in the blocks above can be added, see also Biau and Cadre [3].

The random variables $Z'_{\mathbf{i},\mathbf{n},x}$ are now into blocks of different sizes. Let

$$U(1, \mathbf{n}, x, \mathbf{j}) = \sum_{i_k = 2j_kp+1, \; 1 \leq k \leq N}^{(2j_k+1)p} Z'_{\mathbf{i},\mathbf{n},x},$$

$$U(2, \mathbf{n}, x, \mathbf{j}) = \sum_{i_k = 2j_kp+1, \; 1 \leq k \leq N-1}^{(2j_k+1)p} \; \sum_{i_N = (2j_N+1)p+1}^{2(j_N+1)p} Z'_{\mathbf{i},\mathbf{n},x},$$

$$U(3, \mathbf{n}, x, \mathbf{j}) = \sum_{i_k = 2j_kp+1, \; 1 \leq k \leq N-2}^{(2j_k+1)p} \; \sum_{i_{N-1} = (2j_{N-1}+1)p+1}^{2(j_{N-1}+1)p} \; \sum_{i_N = 2j_Np+1}^{(2j_N+1)p} Z'_{\mathbf{i},\mathbf{n},x},$$

$$U(4, \mathbf{n}, x, \mathbf{j}) = \sum_{i_k = 2j_kp+1, \; 1 \leq k \leq N-2}^{(2j_k+1)p} \; \sum_{i_{N-1} = (2j_{N-1}+1)p+1}^{2(j_{N-1}+1)p} \; \sum_{i_N = (2j_N+1)p+1}^{2(j_N+1)p} Z'_{\mathbf{i},\mathbf{n},x},$$

and so on. Note that

$$U(2^{N-1}, \mathbf{n}, x, \mathbf{j}) = \sum_{i_k = (2j_k+1)p+1, \; 1 \leq k \leq N-1}^{2(j_k+1)p} \; \sum_{i_N = 2j_Np+1}^{(2j_N+1)p} Z'_{\mathbf{i},\mathbf{n},x}.$$

Finally,

$$U(2^N, \mathbf{n}, x, \mathbf{j}) = \sum_{i_k = (2j_k+1)p+1, \; 1 \leq k \leq N}^{2(j_k+1)p} Z'_{\mathbf{i},\mathbf{n},x}.$$

Setting $\mathcal{T} = \{0, ..., t_1 - 1\} \times ... \times \{0, ..., t_N - 1\}$, we define for each integer $l = 1, ..., 2^N$,

$$T(\mathbf{n}, x, l) = \sum_{\mathbf{j} \in \mathcal{T}} U(l, \mathbf{n}, x, \mathbf{j}).$$

Then we obtain the following decomposition

$$Q_{\mathbf{n}}(x) = \varphi_{\mathbf{n}}(x) - E\varphi_{\mathbf{n}}(x) = \sum_{l=1}^{2^N} T(\mathbf{n}, x, l).$$

To prove that $Q_{\mathbf{n}}(x) = O\left(\sqrt{\frac{\log \widehat{\mathbf{n}}}{\widehat{\mathbf{n}} p^x_{h_{\mathbf{n}}}}}\right)$ in probability or *a.s.*, it is sufficient to show that

$$T(\mathbf{n}, x, l) = O\left(\sqrt{\frac{\log \widehat{\mathbf{n}}}{\widehat{\mathbf{n}} p^x_{h_{\mathbf{n}}}}}\right) \text{ a.s. or in probability} \tag{6.6}$$

for each $l = 1, ..., N$. Without loss of generality we will show (6.6) for $l = 1$.
Let us prove that given an arbitrary large positive constant c, there exists a positive constant C such that for any $\eta > 0$

$$P\left[|T(\mathbf{n}, x, 1)| > \eta\sqrt{\frac{\log \widehat{\mathbf{n}}}{\widehat{\mathbf{n}} p^x_{h_{\mathbf{n}}}}}\right] \le C(\widehat{\mathbf{n}}^{-c} + \beta_{1\widehat{\mathbf{n}}})$$

where $\beta_{1\widehat{\mathbf{n}}} = \chi(\widehat{\mathbf{n}}, p^N)\varphi(p)\epsilon_{\mathbf{n}}^{-1}$.
Let

$$T(\mathbf{n}, x, 1) = \sum_{\mathbf{j} \in \mathcal{T}} U(1, \mathbf{n}, x, \mathbf{j}).$$

be the sum of

$$\widehat{\mathbf{t}} = t_1 \times ... \times t_N \tag{6.7}$$

of the $U(1, \mathbf{n}, x, \mathbf{j})$'s. Note that $U(1, \mathbf{n}, x, \mathbf{j})$ is measurable with respect to the $\sigma-$field generated by $X_{\mathbf{i}}$ with $\mathbf{i}$ belonging to the set of sites

$$\mathcal{I}_{\mathbf{i},\ \mathbf{j}} = \{\mathbf{i} :\ 2j_k p + 1 \le i_k \le (2j_k + 1)p,\ k = 1, ..., N\}$$

These sets of sites are separated by a distance greater than p. Enumerate the random variables's $U(1, \mathbf{n}, x, \mathbf{j})$ and the corresponding $\sigma-$field with which they are measurable in an arbitrary manner and refer to them respectively as $V_1, ..., V_{\widehat{\mathbf{t}}}$ and $\mathcal{B}_1, ..., \mathcal{B}_{\widehat{\mathbf{t}}}$. By Lemma 6.1 of Carbon et al (1997), we approximate $V_1, ..., V_{\widehat{\mathbf{t}}}$ by $V_1^*, ..., V_{\widehat{\mathbf{t}}}^*$. We have

$$|V_i| = |U(1, \mathbf{n}, x, \mathbf{j})| < C p^N \widehat{\mathbf{n}}^{-1}. \tag{6.8}$$

Let $\epsilon_{\mathbf{n}} = \eta\sqrt{\frac{\log \widehat{\mathbf{n}}}{\widehat{\mathbf{n}} p^x_{h_{\mathbf{n}}}}}$ where $\eta > 0$ is a constant to be chosen later.

Since we have $T(\mathbf{n}, x, 1) = \sum_{i=1}^{\widehat{\mathbf{t}}} V_i$, then

$$P\left[|T(\mathbf{n}, x, 1)| > \epsilon_{\mathbf{n}}\right] \le P\left[\left|\sum_{i=1}^{\widehat{\mathbf{t}}} V_i^*\right| > \epsilon_{\mathbf{n}}/2\right] + P\left[\sum_{i=1}^{\widehat{\mathbf{t}}} |V_i - V_i^*| > \epsilon_{\mathbf{n}}/2\right]. \tag{6.9}$$

By Markov's inequality and using (6.8), (6.1) and recall that the sets of sites with respect to which V_i's are measurable are separated by a distance greater than p, we get

$$P\left[\sum_{i=1}^{\widehat{\mathbf{t}}} |V_i - V_i^*| > \epsilon_{\mathbf{n}}\right] \leq C\widehat{\mathbf{t}}p^N \widehat{\mathbf{n}}^{-1}\chi(\widehat{\mathbf{n}}, p^N)\psi(p)\epsilon_{\mathbf{n}}^{-1} \sim \beta_{1\widehat{\mathbf{n}}}. \tag{6.10}$$

Let

$$\lambda_{\mathbf{n}} = \left(\widehat{\mathbf{n}} p_{h_{\mathbf{n}}}^x \log \widehat{\mathbf{n}}\right)^{1/2}, \tag{6.11}$$

$$p = \left[\left(\frac{\widehat{\mathbf{n}} p_{h_{\mathbf{n}}}^x}{4\lambda_{\mathbf{n}}}\right)^{1/N}\right] \sim \left(\frac{\widehat{\mathbf{n}} p_{h_{\mathbf{n}}}^x}{\log \widehat{\mathbf{n}}}\right)^{1/2N}. \tag{6.12}$$

It is clear that, $\lambda_{\mathbf{n}}\epsilon_{\mathbf{n}} = \eta \log \widehat{\mathbf{n}}$.

If (2.5) holds for $\theta > 2N$, we have (by Lemma 3)

$$\lim_{\mathbf{n}\to\infty} \widehat{\mathbf{n}} p_{h_{\mathbf{n}}}^x (R_{\mathbf{n}(x)} + U_{\mathbf{n}(x)}) < C \tag{6.13}$$

where C is a positive constant and

$$U_{\mathbf{n}(x)} = \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} E(Z_{\mathbf{i},\mathbf{n},x})^2$$

$$R_{\mathbf{n}(x)} = \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} \sum_{\mathbf{l}\in\mathcal{I}_{\mathbf{n}},\ i_k\neq l_k \text{ for some } k} |Cov(Z_{\mathbf{i},\mathbf{n},x},\ Z_{\mathbf{l},\mathbf{n},x})|.$$

By (6.3) and (6.13), we have

$$\lambda_{\mathbf{n}}^2 \sum_{i=1}^{\widehat{\mathbf{t}}} E(V_i^*)^2 \leq C\widehat{\mathbf{n}} p_{h_{\mathbf{n}}}^x (U_{\mathbf{n}(x)} + R_{\mathbf{n}(x)}) \log \widehat{\mathbf{n}} < C \log \widehat{\mathbf{n}}.$$

Using (6.8), we get $|\lambda_{\mathbf{n}} V_i^*| < 1/2$ for large $\widehat{\mathbf{n}}$. We deduce from Berstein's inequality that

$$P\left[\left|\sum_{i=1}^{\widehat{\mathbf{t}}} V_i^*\right| > \epsilon_{\mathbf{n}}\right] \leq 2\exp(-\lambda_{\mathbf{n}}\epsilon_{\mathbf{n}} + \lambda_{\mathbf{n}}^2 \sum_{i=1}^{\widehat{\mathbf{t}}} E(V_i^*)^2) \leq 2\exp((-\eta + C)\log \widehat{\mathbf{n}}) \leq \widehat{\mathbf{n}}^{-c} \tag{6.14}$$

for sufficiently large $\widehat{\mathbf{n}}$. We deduce from (6.9), (6.10) and (6.14) that

$$P[|T(\mathbf{n}, x, 1)| > \epsilon_{\mathbf{n}}] \leq C(\widehat{\mathbf{n}}^{-c} + \beta_{1\widehat{\mathbf{n}}})$$

To complete the proof, we will show that $\beta_{1\widehat{\mathbf{n}}} \to 0$. Recall that

$$\beta_{1\widehat{\mathbf{n}}} = \chi(\widehat{\mathbf{n}}, p^N)\psi(p)\epsilon_{\mathbf{n}}^{-1}. \tag{6.15}$$

Furthermore, condition (2.3) allows to show that $\theta > 2N$ is equivalent to $(\beta_{1\widehat{\mathbf{n}}})^{-1} \to \infty$. Actually, we have

$$\beta_{1\widehat{\mathbf{n}}} \leq C\chi(\widehat{\mathbf{n}}, p^N)p^{-\theta}\epsilon_{\mathbf{n}}^{-1}$$

$$\leq C\left(\frac{\widehat{\mathbf{n}}p^x_{h_{\mathbf{n}}}}{\log\widehat{\mathbf{n}}}\right)^{(1/2)-(\theta/2N)} \left((\widehat{\mathbf{n}}p^x_{h_{\mathbf{n}}}/\log\widehat{\mathbf{n}})\right)^{\frac{1}{2}}$$

$$= C\left(\frac{\widehat{\mathbf{n}}p^x_{h_{\mathbf{n}}}}{\log\widehat{\mathbf{n}}}\right)^{\frac{-\theta+2N}{2N}}$$

Analogously to the condition (2.3), we have under (2.4):

$$\beta_{1\widehat{\mathbf{n}}} \leq C\psi(\widehat{\mathbf{n}}, p^N)p^{-\theta}\epsilon_{\mathbf{n}}^{-1}$$

$$\leq C\widehat{\mathbf{n}}^{\widetilde{\beta}}\left(\frac{\widehat{\mathbf{n}}p^x_{h_{\mathbf{n}}}}{\log\widehat{\mathbf{n}}}\right)^{(1/2)-(\theta/2N)}$$

$$= C\left(\widehat{\mathbf{n}}(p^x_{h_{\mathbf{n}}}/\log\widehat{\mathbf{n}})^{\theta_1}\right)^{\frac{-\theta+N(2\widetilde{\beta}+1)}{2N}}.$$

Then, condition $\left(\widehat{\mathbf{n}}(p^x_{h_{\mathbf{n}}}/\log\widehat{\mathbf{n}})^{\theta_1}\right) \to \infty$ leads to $\beta_{1\widehat{\mathbf{n}}} \to 0$. ▬

Proof. of Theorem 5

Under (2.3), we get (see the proof of Theorem 4):

$$\beta_{1\widehat{\mathbf{n}}} \leq C\left(\frac{\widehat{\mathbf{n}}p^x_{h_{\mathbf{n}}}}{\log\widehat{\mathbf{n}}}\right)^{\frac{-\theta+2N}{2N}},$$

then

$$\beta_{1\widehat{\mathbf{n}}}\widehat{\mathbf{n}}g(\mathbf{n}) \leq \left(\widehat{\mathbf{n}}\left(\frac{p^x_{h_{\mathbf{n}}}}{\log\widehat{\mathbf{n}}}\right)^{\theta_1^*} g(\mathbf{n})^{\frac{2N}{4N-\theta}}\right)^{\frac{4N-\theta}{2N}}.$$

Analogously to (2.3), we have under (2.4):

$$\beta_{1\widehat{\mathbf{n}}}\widehat{\mathbf{n}}g(\mathbf{n}) \leq \left(\widehat{\mathbf{n}}\left(\frac{p^x_{h_{\mathbf{n}}}}{\log\widehat{\mathbf{n}}}\right)^{\theta_2^*} g(\mathbf{n})^{\frac{2N}{N(2\widetilde{\beta}+3)-\theta}}\right)^{\frac{N(2\widetilde{\beta}+3)-\theta}{2N}}$$

Then the assumption $\left(\widehat{\mathbf{n}}\left(\frac{p^x_{h_{\mathbf{n}}}}{\log\widehat{\mathbf{n}}}\right)^{\theta_1^*} g(\mathbf{n})^{\frac{2N}{4N-\theta}}\right)^{\frac{4N-\theta}{2N}} \to \infty$ or $\left(\widehat{\mathbf{n}}\left(\frac{p^x_{h_{\mathbf{n}}}}{\log\widehat{\mathbf{n}}}\right)^{\theta_2^*} g(\mathbf{n})^{\frac{2N}{N(2\widetilde{\beta}+3)-\theta}}\right)^{\frac{N(2\widetilde{\beta}+3)-\theta}{2N}} \to \infty$ implies that $\beta_{1\widehat{\mathbf{n}}}\widehat{\mathbf{n}}g(\mathbf{n}) \to 0$, this last yields that $\sum_{\mathbf{n}\in\mathbb{N}^N}\beta_{1\widehat{\mathbf{n}}} < \infty$ and the proof is complete by Borel-Cantelli's Lemma. ▬

6.2. Proofs of results of uniform convergence case.

Proof. of Lemma 6: It is similar to that of Lemma 3, it suffices to replace $p^x_{h_{\mathbf{n}}}$ by $\Gamma(h_{\mathbf{n}})$. ■

To show theorems 7-10, we consider the fact that the set $\mathcal{C}$ is covered by $d_{\mathbf{n}}$ balls $B_k = B(x_k,\, \rho_{\mathbf{n}})$ of radius $\rho_{\mathbf{n}}$ and center at x_k and we define:

$$S_{1\mathbf{n}} = \max_{1\le k\le d_{\mathbf{n}}} \sup_{x\in B_k} |\varphi_{\mathbf{n}}(x) - \varphi_{\mathbf{n}}(x_k)|,$$

$$S_{2\mathbf{n}} = \max_{1\le k\le d_{\mathbf{n}}} \sup_{x\in B_k} |E\varphi_{\mathbf{n}}(x_k) - E\varphi_{\mathbf{n}}(x)|,$$

$$S_{3\mathbf{n}} = \max_{1\le k\le d_{\mathbf{n}}} |\varphi_{\mathbf{n}}(x_k) - E\varphi_{\mathbf{n}}(x_k)|.$$

Then,

$$\sup_{x\in\mathcal{C}} |\varphi_{\mathbf{n}}(x) - E\varphi_{\mathbf{n}}(x)| \le S_{1\mathbf{n}} + S_{2\mathbf{n}} + S_{3\mathbf{n}}.$$

It easy to see that under H_2 and HK_2, $S_{1\mathbf{n}}$ and $S_{2\mathbf{n}}$ are equal to $o\left(\sqrt{\frac{\log\widehat{\mathbf{n}}}{\Gamma(h_{\mathbf{n}})\widehat{\mathbf{n}}}}\right)$ in probability (or almost surely).

It remains to show that $S_{3\mathbf{n}} = O\left(\sqrt{\frac{\log\widehat{\mathbf{n}}}{\Gamma(h_{\mathbf{n}})\widehat{\mathbf{n}}}}\right)$ in probability (resp. almost surely) which is equivalent to show that $\max_{1\le j\le d_{\mathbf{n}}} |T(\mathbf{n}, x_j, 1)| = O\left(\sqrt{\frac{\log\widehat{\mathbf{n}}}{\Gamma(h_{\mathbf{n}})\widehat{\mathbf{n}}}}\right)$ probability (resp. almost surely).

Proof. of Theorem 7

We have:

$$P[\sup_{x\in\mathcal{C}} |T(\mathbf{n}, x, 1)| > \epsilon_{\mathbf{n}}] \le C d_{\mathbf{n}}(\widehat{\mathbf{n}}^{-c} + \beta_{1\widehat{\mathbf{n}}}),$$

where $p^x_{h_{\mathbf{n}}}$ is replaced by $\Gamma(h_{\mathbf{n}})$ in the expressions of $\epsilon_{\mathbf{n}}$ and $\beta_{1\widehat{\mathbf{n}}}$ of the previous proof. To complete the proof, we will show that $d_{\mathbf{n}}\widehat{\mathbf{n}}^{-c} \to 0$ and $d_{\mathbf{n}}\beta_{1\widehat{\mathbf{n}}} \to 0$. We have

$$d_{\mathbf{n}}\widehat{\mathbf{n}}^{-c} \le C\widehat{\mathbf{n}}^{\beta-c}$$

which goes to zero if $c > \beta$. We get the following inequality, under (2.3):

$$d_{\mathbf{n}}\beta_{1\widehat{\mathbf{n}}} \le C\widehat{\mathbf{n}}^{\beta}\left(\frac{\widehat{\mathbf{n}}\Gamma(h_{\mathbf{n}})}{\log\widehat{\mathbf{n}}}\right)^{\frac{-\theta+2N}{2N}} = \left(\widehat{\mathbf{n}}(\Gamma(h_{\mathbf{n}})/\log\widehat{\mathbf{n}})^{\theta_2}\right)^{\frac{-\theta+2N(\beta+1)}{2N}}. \qquad (6.16)$$

Analogously to (2.3), we have under (2.4):

$$d_{\mathbf{n}}\beta_{1\widehat{\mathbf{n}}} \leq C\widehat{\mathbf{n}}^{\beta}\widehat{\mathbf{n}}^{\widetilde{\beta}}\left(\frac{\widehat{\mathbf{n}}\Gamma(h_{\mathbf{n}})}{\log\widehat{\mathbf{n}}}\right)^{(1/2)-(\theta/2N)} = \left(\widehat{\mathbf{n}}(\Gamma(h_{\mathbf{n}})/\log\widehat{\mathbf{n}})^{\theta_3}\right)^{\frac{-\theta+N(2\beta+2\widetilde{\beta}+1)}{2N}}. \tag{6.17}$$

This yields the proof. ■

Proof. of Theorem 8

We have under (2.3):

$$d_{\mathbf{n}}\beta_{1\widehat{\mathbf{n}}} \leq C\widehat{\mathbf{n}}^{\beta}\left(\frac{\widehat{\mathbf{n}}\Gamma(h_{\mathbf{n}})}{\log\widehat{\mathbf{n}}}\right)^{\frac{-\theta+2N}{2N}},$$

Then,

$$d_{\mathbf{n}}\beta_{1\widehat{\mathbf{n}}}\widehat{\mathbf{n}}g(\mathbf{n}) \leq C\left(\widehat{\mathbf{n}}\left(\frac{\Gamma(h_{\mathbf{n}})}{\log\widehat{\mathbf{n}}}\right)^{\theta_3^*}g(\mathbf{n})^{\frac{2N}{N(2\beta+4)-\theta}}\right)^{\frac{N(2\beta+4)-\theta}{2N}}.$$

In the case where (2.4) is satisfied, we get:

$$d_{\mathbf{n}}\beta_{1\widehat{\mathbf{n}}}\widehat{\mathbf{n}}g(\mathbf{n}) \leq C\left(\widehat{\mathbf{n}}\left(\frac{\Gamma(h_{\mathbf{n}})}{\log\widehat{\mathbf{n}}}\right)^{\theta_4^*}g(\mathbf{n})^{\frac{2N}{N(2\beta+2\widetilde{\beta}+3)-\theta}}\right)^{\frac{N(2\beta+2\widetilde{\beta}+3)-\theta}{2N}}.$$

Then, condition $\left(\widehat{\mathbf{n}}(\Gamma(h_{\mathbf{n}})/\log\widehat{\mathbf{n}})^{\theta_3^*}(g(\mathbf{n}))^{\frac{2N}{-\theta+2N(\beta+2)}}\right) \to \infty$ or $\left(\widehat{\mathbf{n}}(\Gamma(h_{\mathbf{n}})/\log\widehat{\mathbf{n}})^{\theta_4^*}(g(\mathbf{n}))^{\frac{2N}{-\theta+N(2\beta+2\widetilde{\beta}+3)}}\right) \to \infty$ is equivalent to $\widehat{\mathbf{n}}g(\mathbf{n})\beta_{1\widehat{\mathbf{n}}} \to 0$ which implies $\sum_{\mathbf{n}\in\mathbb{N}^N}\beta_{1\widehat{\mathbf{n}}} < \infty$, then the theorem follows by Borel-Cantelli Lemma. ■

Proof. of Theorem 9

According to Theorem 7, we have:

$$\sup_{x\in\mathcal{C}}|\varphi_{\mathbf{n}}(x) - E(\varphi_{\mathbf{n}}(x))| = O\left(\sqrt{\frac{\log\widehat{\mathbf{n}}}{\Gamma(h_{\mathbf{n}})\widehat{\mathbf{n}}}}\right), \quad \text{in probability.}$$

and

$$\sup_{x\in\mathcal{C}}|f_{\mathbf{n}}(x) - 1| = O\left(\sqrt{\frac{\log\widehat{\mathbf{n}}}{\Gamma(h_{\mathbf{n}})\widehat{\mathbf{n}}}}\right), \quad \text{in probability.}$$

We get also by HF_3 (see the proof of Theorem 4):

$$\sup_{x\in\mathcal{C}} |r(x) - E(\varphi_{\mathbf{n}}(x))| \leq \sup_{x\in\mathcal{C}} \sup_{u\in B(x,h_{\mathbf{n}})} |r(x) - r(u)| = O\left(h_{\mathbf{n}}\right).$$

$\blacksquare$

Proof. of Theorem 10

According to Theorem 8 and HF_3, we have:

$$\sup_{x\in\mathcal{C}} |\varphi_{\mathbf{n}}(x) - E(\varphi_{\mathbf{n}}(x))| = \sup_{x\in\mathcal{C}} |f_{\mathbf{n}}(x) - 1| = O\left(\sqrt{\frac{\log \widehat{\mathbf{n}}}{\Gamma(h_{\mathbf{n}})\widehat{\mathbf{n}}}}\right), \quad \text{a.s}$$

and

$$\sup_{x\in\mathcal{C}} |r(x) - E(\varphi_{\mathbf{n}}(x))| \leq \sup_{x\in\mathcal{C}} \sup_{u\in B(x,h_{\mathbf{n}})} |r(x) - r(u)| = O\left(h_{\mathbf{n}}\right).$$

The proof is therefore complete. $\blacksquare$

References

[1] Anselin, L., Florax, R. J. M., 1995. New Directions in Spatial Econometrics. Springer, Berlin.

[2] Besse, P., Cardot, H., Stephenson, D., 2000. Autoregressive forecasting of some functional climatic variations. Scand. J. Statist. 27, 673–687.

[3] Biau, G., Cadre, B., 2004. Nonparametric spatial prediction. Stat. Inference Stoch. Process. 7, 327–349.

[4] Bosq, D., 2000. Linear Processes in Function Spaces: Theory and applications. Lecture Notes in Statistics, Springer-Verlag, New York.

[5] Carbon, M., Francq, C., Tran, L. T., 2007. Kernel regression estimation for random fields. J. Statist. Plann. Inference 137 (Issue 3), 778–798.

[6] Carbon, M., Hallin, M., Tran, L. T., 1996. Kernel density estimation for random fields. Stat. Proba. Lett. 36, 115–125.

[7] Carbon, M., Tran, L. T., Wu, B., 1997. Kernel density estimation for random fields: The l_1 theory. J. Nonparametr. Stat. 6, 157–170.

[8] Cardot, H., Johannes, J., 2010. Thresholding projection estimators in functional linear models. J. Multivariate Anal. 101, 395–408.

[9] Chilès, J., P., D., 1999. Geostatistics. Modeling spatial uncertainty. Wiley Series in Probability and Mathematical Statistics, New-York.

[10] Cressie, N. A., 1991. Statistics for Spatial Data. Wiley Series in Probability and Mathematical Statistics, New-York.

[11] Dabo-Niang, S., Ferraty, F., 2008. Functional and Operatorial Statistics. Springer-Verlag, New-York.

[12] Dabo-Niang, S., Rhomari, N., 2009. Kernel regression estimate in a banach space. J. Statist. Plann. Inf. 139, 1421–1434.

[13] Dabo-Niang, S., Yao, A., Pischedda, L., Cuny, P., Gilbert, F., 2009. Spatial kernel mode estimation for functional random, with application to bioturbation problem. Stochastic Environmental Research and Risk Assessment 24 (4), 487–497.

[14] Dabo-Niang, S., Yao, A. F., 2007. Kernel regression estimation for continuous spatial processes. Math. Method. Stat. 16 (4), 1–20.

[15] Delicado, P., Giraldo, R., Mateu, J., 2008, pp 135-141. Point-wise kriging for spatail prediction of functional data. Springer Verlag, Functional and Operatorial Statistics, edited by S. Dabo-Niang and F. Ferraty.

[16] Doukhan, P., 1994. Mixing - Properties and Examples. Lecture Notes in Statistics, Springer-Verlag, New York.

[17] Fernandez de Castro, B., Guillas, S., W., G. M., 2005. Technometrics 47 (2), 212–222.

[18] Ferraty, F., Vieu, P., 2006. Nonparametric Functional Data Analysis. Springer-Verlag, berlin.

[19] Ferré, L., Yao, A., 2005. Smoothed functional inverse regression. Statist. Sinica 15 (3), 665–683.

[20] Guyon, X., 1995. Random Fields on a Network - Modeling, Statistics, and Applications. Springer, New-York.

[21] Hallin, M., L. Z., Yu, K., 2009. Local linear spatial quantile regression. Bernoulli 15, 659–686.

[22] Hallin, M., Lu, Z., Tran, L. T., 2004. Local linear spatial regression. Ann. Statist. 32 (6), 2469–2500.

[23] Lakasaci, A., Fouzia, M., 2009. Nonparametric estimation of conditional quantiles for functional and spatial dependent variables. C. R. Acad. Sci. Paris Ser. I Math 347 (17-18), 1075–1080.

[24] Li, J., Tran, L., 2008. Nonparametric estimation of conditional expectation. Statist. Plann. Inference 139, 164–175.

[25] Lu, Z., Chen, X., 2002. Spatial nonparametric regression estimation: Non-isotropic case. Acta Math. Appl. Sin. Engl. Ser. 18 (4), 641–656.

[26] Lu, Z., Chen, X., 2004. Spatial kernel regression estimation: Weak concistency. Stat. Proba. Lett 68, 125–136.

[27] Nadaraya, E., 1964. On estimating regression. Theor.Prob.Appl., 141–142.

[28] Nerini, D., Monestiez, P., Manté, C., 2010. A cokriging method for spatial functional. Journal of Multivariate Analysis 101 (2), 409–418.

[29] Rachdi, M., Vieu, P., 2007. Nonparametric regression for functional data: automatic smoothing parametrer selection. Statist. Plann. Inference 137 (9), 2784–2801.

[30] Ramsay, J., Silverman, B., 2002. Applied functional data analysis; Methods and case studies. Springer-Verlag, New York.

[31] Rio, E., 2000. Théorie Asymptotic des Processus Aléatoires Faiblement Dépendants. Springer-Verlag, Berlin.

[32] Ripley, B., 1981. Spatial Statistics. Wiley, New-York.

[33] Tran, L. T., 1990. Kernel density estimation on random fields. J. Multivariate Anal. 34, 37–53.

[34] Watson, G., 1964. Smooth regression analysis. Sankhya, Ser. A 26, 359–372.